Basic Properties of a time series data

Rishman Jot Kaur Chahal

Assistant Professor
Department of Humanities and Social Sciences
Joint Faculty at the Mehta Family School of DS & AI
Indian Institute of Technology Roorkee

October 5, 2023

Outline

- Definition & Visualization of time series
- Univariate time series models
- Moving Average Process (MA)
- Autoregressive Process (AR)
- ACF & PACF plots
- ARMA Process - Box Jenkins Approach
- ARIMA Models
- Stationarity and Unit root testing

Univariate
Multivariate

Testing

Time Series - definition

- Time series is a series of data points indexed (or listed or graphed) in time order.
- So is the following data is a time series?

say

Given Time (2020)

City (i)	Temperature (Y_i)
Miami, FL	83
Houston, TX	82
Mobile, AL	85
San Francisco, CA	65
New York, NY	76
Boston, MA	72

Y_t	t
83	2020
82	2021
74	2022.
66	

Time Series - definition

- No, it is not as it a cross-sectional data. It is the data at one point in time or time is not considered a factor.
- Example of a time series is as follows:

i = Boston

Time (t)	Temperature (Y_t)
10:00 AM	72
10:05 AM	76
10:10 AM	83
10:15 AM	87
10:20 AM	91
10:25 AM	99
10:30 AM	104
10:35 AM	107
10:40 AM	110
10:45 AM	123
10:50 AM	125

i	T	Temp.
H	10:00	–
⋮	⋮	–
H	10:50	–
B.	10	–
	⋮	–
	10:50	–

Panel or Longitudinal Data

i = H, B, M, ...

t = 10, 10:05, 10:10, ...

Y_{it}

Type of Data?

Time Series forecasting - definition

$Y_t = \text{Temp}_t$ or $Y_t = \text{Sales}_t$.

- Time Series forecasting is the use of a model to predict future values based on previously observed values.

$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots)$

- Example: Based on previous year sales performance of XYZ ltd. we can predict the future sales forecast.

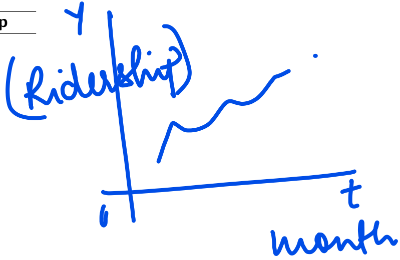
Year (t)	Sales (Y_t).
2010	72000
2011	76000
2012	83000
2013	87000
2014	91000
2015	99000
2016	**104000**
2017	**107000**
2018	**110000**

2010-15.

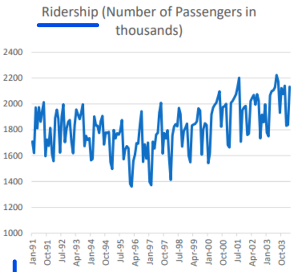
2016-18.

- The most basic and informative plot for visualizing a time series is the **time plot**.
- It is a line chart of the series values $(y_1, y_2, ...)$ over time $(t = 1, 2, ...)$, with temporal labels (like date) on the horizontal axis.

Month	Ridership
Jan-91	1709
Feb-91	1621
Mar-91	1973
Apr-91	1812
May-91	1975
Jun-91	1862
Jul-09	1940
Aug-91	2013
Sep-91	1596
Oct-91	1725
Nov-91	1676
Dec-91	1814
Jan-92	1615
.	.
.	.
.	.
Apr-94	2100

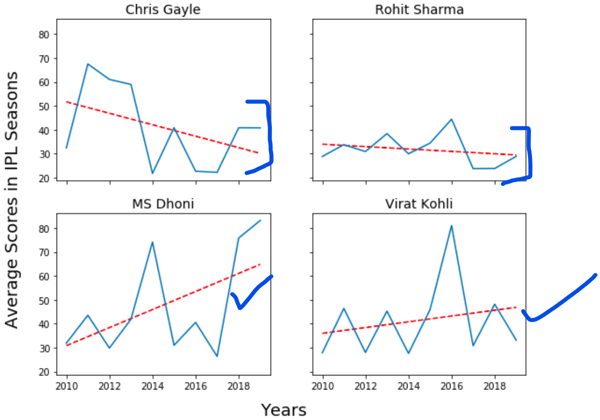


- **Zooming in** is used to reduce the period within the series which can reveal the hidden patterns in the entire series.



Patterns || Seasonality.

- **Adding Trend Line** to further understand the shape of the trend which further could be linear, exponential depending on the type of trend in the data.



- Further, **deseasonalising or suppressing seasonality** provides a better visualization of the data.
- Let's say the sale of ice-cream is higher in summers as compared to winters thus we can look at the overall yearly sale of the ice-cream. This is one way to remove seasonal trend and we can see the overall ice-cream sales.

summers
Winters

- Another way is to plot separate time plots for each season.
- Finally, a popular method is to use **Moving Average** plots.

(MA) Process

- **Lag Scatter Plot**

$Y_t = f(Y_{t-1})$

  - Usually in a time series there exists a relationship between current and previous years observations which are know as lags.
  - A lag plot contains Y_t for all t on vertical axis and Y_{t-k} for all t where k is lag value.

$K = 1, 2, 3 \ldots N$

$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \epsilon_t$

  - Lag plots usually help to understand the distribution of the model i.e. to know the shape of the data.
    - So, if the plot is linear, then the underlying structure is of the autoregressive model. (AR)

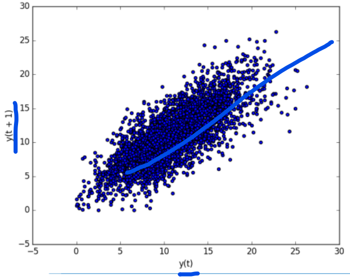
AR(1)

    - And if the lag plot is of elliptical shape, then the underlying structure represents a continuous periodic function such as sine, cosine, etc.
    - Further to know the outliers and seasonality of the data.

$\text{corr.}(\epsilon_t, \epsilon_{t-1}) \neq 0$

    - To understand the failure of a CLRM assumption i.e. **Autocorrelation**. If there is a linear plot it means there exists autocorrelation in the data.

Y_t



Y_{t-1}.

Next, let us look at the time series visualization in python.

- **Lag Scatter Plot - Implementation (advance)**
  - Important Libraries

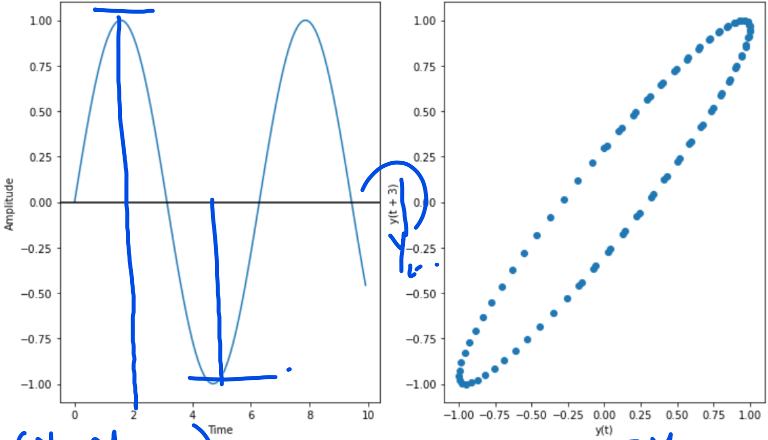
```
#Important Libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats as sc
```

```
# Sine graph and lag plot
time= np.arange(0, 10, 0.1);
amplitude=np.sin(time)
fig, ax = plt.subplots(1, 2, figsize=(12, 7))
ax[0].plot(time, amplitude)
ax[0].set_xlabel('Time')
ax[0].set_ylabel('Amplitude')
ax[0].axhline(y=0, color='k')
amplitude_series = pd.Series(amplitude)
pd.plotting.lag_plot(amplitude_series, lag= 5, ax =ax[1])
plt.show()
```

- **Sine Plot (elliptic curve)**

Y_t.



$Corr(Y_t, Y_{t-3}) = \_\_\_$

$\therefore Y_{t-3}$.

Figure: Sine Plot (elliptic curve)

- **Strong Autocorrelation**

```
# Google Stock and Lag Plot (Strong Autocorrelation)
google_stock_data = pd.read_csv('GOOG.csv')
google_stock_data.reset_index(inplace=True)
fig, ax = plt.subplots(1, 2, figsize=(12, 7))
ax[0].plot(google_stock_data['Adj Close'], google_stock_data['index'])
pd.plotting.lag_plot(google_stock_data['Adj Close'], lag=1,ax=ax[1])
plt.show()
```

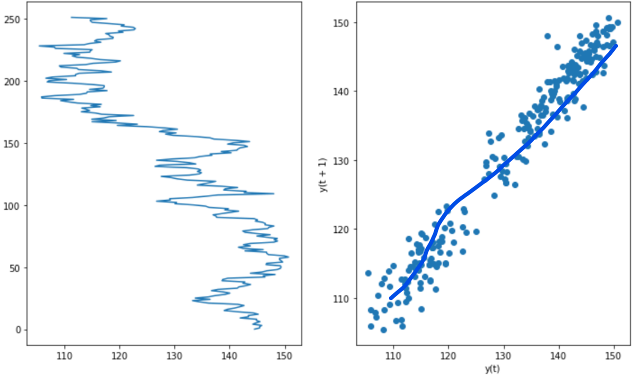


Figure: Google Stock Data (Strong Autocorrelation)

- These models are a class of specifications where one attempts to model and to predict (dependent) variable (say y_t) using only information contained in their own past values (y_{t-1}).
- These models may further include the current and past values of the error term.

$$\boxed{Y_t = \beta Y_{t-1} + \epsilon_t + \gamma \epsilon_{t-1}}$$

- Thus, these models are different from *structured models*. ?
- What do we mean by *structured models*? →

$$Y_t = f(X_t) = \alpha_0 + \alpha_1 X_t + \epsilon_t.$$

- Structured models are multivariate in nature and explains the changes in dependent variable based on the movements in the current or past values of other (explanatory) variables. $Y_t = f(Y_{t-1}, Y_{t-2}) \dots$
- Time Series models are more like a-theoretical indicating their construction is not based upon any underlying theoretical model on the behaviour of a variable. $(X_{1t}, X_{2t}, \dots)$
- But they are an attempt to capture the empirically relevant features of the observed data that may have arisen from a variety of different (but unspecified) structured models.

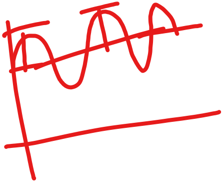
Stationary / Non-stat.

- An important class of time series models is **ARIMA** models (Box and Jenkins, 1976).
- Before that, let us first understand the basic important concepts.

II Order Stationarity

- A Weakly Stationary Process

If a series satisfies the next three equations, it is said to be weakly or covariance stationary.



Mean $$E(y_t) = \mu \quad t = 1, 2, \ldots,$$

Var. $$E(y_t - \mu)(y_t - \mu) = \sigma^2$$

Cov. $$E(y_{t_1} - \mu)(y_{t_2} - \mu) = \gamma_{t_2 - t_1}$$

- A stationary process or series has the following properties:
  -constant mean μ
  -constant variance σ^2.
  -constant autocovariance structure $= \gamma$
- The latter refers to the covariance between y_{t-1} and y_{t-2} being the same as y_{t-5} and y_{t-6}.

$$E(Y_t - \mu)(Y_{t-s} - \mu) = \gamma_s$$

$$s = 0, 1, 2, \ldots$$

- So if the process is stationary, all the variances are the same and all the covariances depend only on the difference between t_1 and t_2. The moments

$$E(y_t - \overset{\mu}{E(y_t)})(y_{t+s} - \overset{\mu}{E(y_{t+s})}) = \gamma_s, \; s = 0, 1, 2, \ldots$$

are known as the covariance function.

- The covariances, γ_s , are also known as autocovariances.
- However, the value of the autocovariances depend on the units of measurement of y_t. (Rs, tonnes, ---)
- It is thus more convenient to use the autocorrelations which are the autocovariances normalised by dividing by the variance:

$\tau_s = \frac{\gamma_s}{\gamma_0}, s = 0, 1, 2, \ldots$

corr x = $\frac{\text{cov}}{\text{var}}$.

- If we plot τ_s against s=0,1,2,... then we obtain the autocorrelation function (acf) or correlogram.

Stationary Series

Summary

- Remember, mathematically

$$E(y_t) = \mu$$

$$E(y_t - \mu)^2 = \sigma^2$$

$$E(y_{t1} - \mu)(y_{t2} - \mu) = \gamma_{t2-t1}, \forall t_1, t_2$$

- A white noise process is one with (virtually) no noticeable structure. A definition of a white noise process is

$$E(y_t) = \mu$$

$$Var(y_t) = \sigma^2$$

Autocov.

$$\gamma_{t-r} = \begin{cases} \sigma^2, & \text{if } t = r \\ 0, & \text{otherwise} \end{cases}$$

$$\Omega \text{ (Var-Cov Matrix)} = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}$$

- Thus, a white noise process has constant mean and variance, and zero autocovariances, except at lag zero.

$$E(Y_{1t}-\mu)(Y_{2t}-\mu) = \gamma_1$$

- Hence, the autocorrelation function will be zero apart from a single peak of 1 at s = 0.

$$\tau_1 = \frac{\gamma_1}{\gamma_0} \qquad = \frac{\gamma_s}{\gamma_0} = \frac{\sigma^2}{\sigma^2} = 1.$$

Ex2.

- If it is further assumed that y_t is distributed normally, then the sample autocorrelation coefficients are also approximately normally distributed.

τ_s Moments → μ_1, μ_2, $Sk = 0$, $K = \frac{\mu_4}{\mu_2^2}$

- τ_s approximately N(0,1/T) where T = sample size and $\hat{\tau}_s$ denotes the autocorrelation coefficient at lag s estimated from a sample.

$\tau_s \sim N(0, \frac{1}{T})$, $T = S.S.$

- We can use this to do significance tests for the autocorrelation τ_s coefficients by constructing a confidence interval.
- For example, a 95% confidence interval would be given by $\pm 0.196 * \frac{1}{\sqrt{T}}$. If the sample autocorrelation coefficient, $\hat{\tau}_s$, falls outside this region for any given value of s, then we reject the null hypothesis that the true value of the coefficient at lag s is zero.

$\hat{\tau}_s$

$H_0 : \tau_s = 0$

- We can also test the joint hypothesis that all m of the τ_k correlation coefficients are simultaneously equal to zero using the Q-statistic developed by Box and Pierce (1970):

Joint

$H_0: \tau_1 = \tau_2 = \tau_3 = \ldots = \tau_m = 0$

- $\boxed{Q = T\sum_{k=1}^{m} \tau_k^2}$

(5) lags in your model

where T = sample size, m = maximum lag length.

$\tau_1, \tau_2, \tau_3, \tau_4, \tau_5$

- The Q-statistic is asymptotically distributed as a χ_m^2.
- However, the Box Pierce test has poor small sample properties, so a variant has been developed, called the Ljung-Box statistic:

$$\boxed{Q^* = T(T+2)\sum_{k=1}^{m} \frac{\tau_k^2}{T-k} \sim \chi_m^2}$$

- This statistic is very useful as a portmanteau (general) test of linear dependence in time series

$T = 100$ | $r_1 = .207$, $r_2 = -0.013$, $r_3 = 0.086$, $r_4 = 0.005$, $r_5 = -0.022$.

- <u>Question:</u>
  Suppose that a researcher had estimated the first 5 autocorrelation coefficients using a series of length 100 observations, and found them to be (from 1 to 5): 0.207, -0.013, 0.086, 0.005, -0.022. Test each of the individual coefficient for significance, and use both the BoxPierce and Ljung-Box tests to establish whether they are jointly significant.

$$Q_{BP} = T \sum_{k=1}^{m} r_k^2.$$

$$= 100\left((0.207)^2 + \ldots + (0.022)^2\right)$$

$$= 5.09.$$

$$Q^*_{LB} = T(T+2) \sum \frac{r_k^2}{T-k}$$

$$= 5.26$$

$\tau_1 = 0.207$.

- Solution:
  A coefficient would be significant (rejecting the null hypothesis) if it lies outside (-0.196,+0.196) at the 5% level, so only the first autocorrelation coefficient is significant.

- For joint hypothesis test, $H_0 : \tau_1 = 0, \tau_2 = 0, \tau_3 = 0, \tau_4 = 0, \tau_5 = 0$.
- The test statistics for the Box–Pierce and Ljung–Box tests are given respectively, as

$$Q = 100x(0.207^2 + (-0.013)^2 + 0.086^2 + 0.005^2 + (-0.022)^2)$$

$= \quad < 11.1$ Accept H_0.

- Thus, Q=5.09 and Q^*=5.26 Compared with a tabulated $\chi^2(5) = 11.1$ at the 5% level, so the 5 coefficients are jointly insignificant.

- Moving Average (MA) is the simplest class of time-series model.
- Let u_t (t=1,2,3,...) be a sequence of independently and identically distributed (iid) random variables or white noise process with $E(u_t) = 0$ and $Var(u_t) = \sigma^2$, then $u_t \sim N(0, \sigma^2)$
$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q}$ is a q^{th} order moving average model MA(q).

Autocov.

$r_1 = \frac{\gamma_1}{\gamma_0}.$

- Its properties are
$E(y_t) = \mu$; $Var(y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2$

$E[(y_t - \mu)(y_{t+s} - \mu)] = \gamma_s.$

Covariances

$$\gamma_s = \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + ... + \theta_q\theta_{q-s})\sigma^2, & \text{for } s = 1, 2, ..., q \\ 0, & \text{for s>q} \end{cases}$$

- Thus, MA has a constant mean, constant variance and autocovariances which may be non-zero to lag q and will always be zero thereafter.

$GDP_t = u_t + (-0.5) u_{t-1} + (0.25) u_{t-2}.$

θ_1 θ_2.

OLS

Min $E u_t^2$

- Consider the following MA(2) process:
  $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$
  where u_t is a zero mean white noise process with variance σ^2.

  $E(Y_t)$ & $V(Y_t)$

  - Calculate the mean and variance of y_t

  τ_s

  - Derive the autocorrelation function for this process (i.e. express the autocorrelations, $\tau_1, \tau_2, \ldots$ as functions of the parameters θ_1 and θ_2).

  $\tau_1 = \frac{\gamma_1}{\gamma_0 (\text{var.})}$

  - If $\theta_1 = -0.5$ and $\theta_2 = 0.25$, sketch the acf of y_t.

$E(Y_t) = E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2})$

$= 0$

$\gamma_1 = E(Y_t - \mu)(Y_{t-1} - E(Y_{t-1}))$

$\boxed{\gamma_1 = E(Y_t)(Y_{t-1})}$

- If $E(u_t) = 0$, Then $E(u_{t-i}) = 0 \forall i$.
- Now, taking expectations on both the sides of the above equation. $E(y_t) = E(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}) = E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2}) = 0$
- Now, Variance, $Var(y_t) = E[y_t - E(y_t)][y_t - E(y_t)]$ but $E(y_t) = 0$, so

$$Var(y_t) = E[(y_t)(y_t)]$$
$$= E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})]$$
$$= E[u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2 + \underbrace{crossproducts}_{=0}]$$

No Autocorr.

But E[cross-products]=0 since cov(u_t, u_{t-s})=0 for $s \neq 0$

$E(u_t, u_{t-s})$
$=0$

So $Var(y_t) = \gamma_0 = E[u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2]$
$= \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2$
$= (1 + \theta_1^2 + \theta_2^2)\sigma^2$

(ii) The acf of y_t.
$\gamma_1 = E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$
$= E[y_t][y_{t-1}]$
$= E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3})]$
$= E[(\theta_1 u_{t-1}^2 + \theta_1\theta_2 u_{t-2}^2)]$
$= \theta_1\sigma^2 + \theta_1\theta_2\sigma^2$
$= (\theta_1 + \theta_1\theta_2)\sigma^2$

$\gamma_2 = E(y_t)(y_{t-2}) = \theta_2\sigma^2.$

$\gamma_3 = E(y_t)(y_{t-3})$
$= 0.$

$$\gamma_2 = E[y_t - E(y_t)][y_{t-2} - E(y_{t-2})]$$
$$= E[y_t][y_{t-2}]$$
$$= E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-2} + \theta_1 u_{t-3} + \theta_2 u_{t-4})]$$
$$= E[(\theta_2 u_{t-2}^2)]$$
$$= \theta_2 \sigma^2$$

$$\gamma_3 = E[y_t - E(y_t)][y_{t-3} - E(y_{t-3})]$$
$$= E[y_t][y_{t-3}]$$
$$= E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-3} + \theta_1 u_{t-4} + \theta_2 u_{t-5})]$$
$$= E[(\theta_2 u_{t-3}^2)]$$
$$= 0$$

$s>2$

So $\gamma_s = 0$ for $s>2$. All autocovariances for the MA(2) process will be zero for any lag length, s, greater than 2.

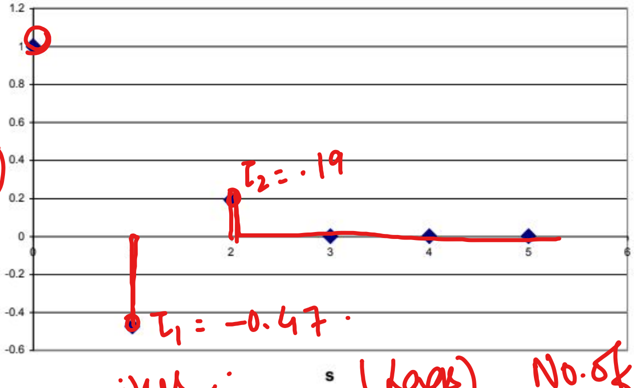
- We have the autocovariances, now calculate the autocorrelations:
- Autocorrelation at lag 0 is:
  $\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$
- Autocorrelation at lag 1 is:
  $\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\theta_1+\theta_1\theta_2)\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{(\theta_1+\theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)}$ — 1
- Autocorrelation at lag 2 is: $\tau_2 = \frac{\gamma_2}{\gamma_0} = \frac{(\theta_2)\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{\theta_2}{(1+\theta_1^2+\theta_2^2)}$ — 2

  $\tau_s = 0 \quad \forall s > 2$
- Thus, $\tau_3 = \frac{\gamma_3}{\gamma_0} = 0$
  $\tau_s = \frac{\gamma_s}{\gamma_0} = 0 \forall s > 2$

(iii) For $\theta_1 = -0.5$ and $\theta_2 = 0.25$, substituting these into the formulae above gives $\tau_1 = -0.476$, $\tau_2 = 0.190$.

Thus the acf plot will appear as follows :

r_s



spikes.

(Lags)

No. of lags that you can include in your MA.

MA(2) → $y_t = \mu + \theta_1 u_{t-1} + \theta_2 u_{t-2}$ (~~y_{t-1}~~, y_{t-2})

- An autoregressive model of order p, an AR(p) can be expressed as
  $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + u_t$ — AR(p)
- Or using the lag operator notation : $Ly_t = y_{t-1}$ $\quad L^i y_t = y_{t-i}$
  $y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$
- or $y_t = \mu + \sum_{i=1}^{p} \phi_i L^i y_t + u_t$
  or $\phi(L) y_t = \mu + u_t$ where $\phi(L) = 1 - (\phi_1 L + \phi_2 L^2 + ... + \phi_p L^p)$

ARIMA (p, q) → Both AR (p) & MA (q).

Stationary (p, 0, q) → I ~ (0)

- Now, setting $\mu = 0$ in the above equation

$$\phi(L)y_t = u_t$$

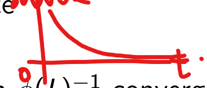
$y_t = (1) y_{t-1} + u_t$

$y_t = L y_t + u_t$

Random Walk

- it would be stated that the process is stationary if it is possible to write

shock

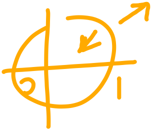


$$y_t = \phi(L)^{-1} u_t$$

with $\phi(L)^{-1}$ converging to zero which means that the autocorrelations will decline eventually as the lag length is increased.

- The condition for stationarity of a general AR(p) model is that the roots of the 'characteristic equation'

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$$



all lie outside the unit circle.

- Why?

- The notion of a characteristic equation is so-called because its roots determine the characteristics of the process y_t. Its like the acf of an AR process will depend on the roots of this characteristic equation which is a polynomial in z.

- **Example 1**: Is $y_t = y_{t-1} + u_t$ stationary?

Unit Root Process

$y_t - y_{t-1} = u_t$.

$y_t(1-L) = u_t \Rightarrow$ $L=1$

$\Rightarrow$ Non-stat.

- The characteristic root is 1, which lies on, not outside, the unit circle, so it is a unit root process (implies non-stationary).

- **Example 2**: Is $y_t = 3y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + u_t$ stationary?

$y_t(1 - 3L + 0.275L^2 - 0.75L^3) = u_t$.

$L = 1/2, 1, 2/3$.

[Coeff. < 1. character. Roots > 1]

The characteristic roots for example 2 are 1, 2/3, and 2. Since only one of these lies outside the unit circle, the process for y_t is non-stationary.

- Consider the following simple AR(1) model
  $y_t = \mu + \phi_1 y_{t-1} + u_t$

$E(Y_t) = \mu + \phi_1 E(Y_{t-1})$

$= \mu + \phi_1 (\mu + \phi_1 E(Y_{t-2}))$

(i) Calculate the (unconditional) mean of y_t.

For the remainder of the question, set $\mu = 0$ for simplicity.

(ii) Calculate the (unconditional) variance of y_t.

$\tau_1 = \frac{\gamma_1}{\gamma_0}$

(iii) Derive the autocorrelation function for y_t.

$\gamma_1 = ?$

$\gamma_1 = Cov(Y_t, Y_{t-1}) = E[(Y_t - \underbrace{E(Y_t)}_{=0})(Y_{t-1} - \underbrace{E(Y_{t-1})}_{=0})]$

(i) Unconditional mean :
$E(y_t) = E(\mu + \phi_1 y_{t-1}$
$= \mu + \phi_1 E(y_{t-1})$
But also

So $E(y_t) = \mu + \phi_1(\mu + \phi_1 E(y_{t-2}))$
$= \mu + \phi_1\mu + \phi_1^2 E(y_{t-2}))$

$E(y_t) = \mu + \phi_1\mu + \phi_1^2 E(y_{t-2}))$
$= \mu + \phi_1\mu + \phi_1^2(\mu + \phi_1 E(y_{t-3}))$
$= \mu + \phi_1\mu + \phi_1^2\mu + \phi_1^3 E(y_{t-3}))$

$n \to \infty$

Making n such substitutions would give
$E(y_t) = \mu(1 + \phi_1 + \phi_1^2 + ... + \phi_1^{n-1}) + \boxed{\phi_1^n E(y_{t-n})}$

$S_n = \frac{a}{(1-r)}$

$S_\infty = \frac{a}{(1-r)}$

An infinite number of such substitutions would give
$E(y_t) = \mu(1 + \phi_1 + \phi_1^2 + \ldots) + \phi_1^\infty y_0 \rightarrow 0$

So long as the model is stationary, i.e. $|\phi_1| < 1$, then $\boxed{\phi_1^\infty = 0}$. $n \rightarrow \infty$
So $E(y_t) = \mu(1 + \phi_1 + \phi_1^2 + \ldots) = \frac{\mu}{1-\phi_1}$

Ly_t.

(ii) Calculating the variance of y_t : $y_t = \phi_1 y_{t-1} + u_t$ — $AR(1)$
$y_t(1 - \phi_1 L) = u_t$ $\mu = 0$
$y_t = (1 - \phi_1 L)^{-1} u_t$

$y_t(1 - \phi_1 L) = u_t$. $y_t - \phi_1 L y_t = u_t$

From Wald's decomposition theorem, $AR(\rho)$ can be expressed as $MA(\infty)$:
$y_t = (1 + \phi_1 L + \phi_1^2 L^2 + \ldots)u_t$ or

$AR(1) \rightarrow MA(\infty)$

$$y_t = u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \ldots$$

OR

$MA(1) \rightarrow AR(\infty)$

So long as $|\phi_1| < 1$, i.e. so long as the process for y_t is stationary, this sum will converge.

$$Var(y_t) = E[y_t - E(y_t)][y_t - E(y_t)]$$

but $E(y_t) = 0$, since we are setting $\mu = 0$.

$Var(y_t) = E[(y_t)(y_t)]$

$= E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + ...)(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + ...)]$

$= E[(u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + ... + \underbrace{cross - products}_{=0.})]$

$= E[(u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + ...)]$

$= \sigma_u^2 + \phi_1^2 \sigma_u^2 + \phi_1^4 \sigma_u^2 + ...$

$= \sigma_u^2 (1 + \phi_1^2 + \phi_1^4 + ...)$ $\rightarrow$ $\underset{n \to \infty}{S_n} = \frac{a}{(1-r)}, \; |r| < 1.$

$= \frac{\sigma_u^2}{(1-\phi_1^2)}$

(provided $|\phi_1| < 1$ we can use the infinite sum)

(iii) Turning now to calculating the acf, first calculate the autocovariances:
$\gamma_1 = cov(y_t, y_{t-1}) = E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$

Since μ has been set to zero, $E(y_t) = 0$ and $E(y_{t-1}) = 0$, so

$$\gamma_1 = \underline{E[(y_t)(y_{t-1})]}$$
$$= E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \ldots)(u_{t-1} + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \ldots)]$$
$$= E[(\phi_1 u_{t-1}^2 + \phi_1^3 u_{t-2}^2 + \ldots + cross - products] \quad = 0$$
$$= \phi_1 \sigma^2 + \phi_1^3 \sigma^2 + \phi_1^5 \sigma^2 + \ldots$$
$$= \left(\frac{\phi_1^2 \sigma^2}{(1-\phi_1^2)}\right)$$

No Auto

$\underline{E(u_{t-1}^2)} = \underline{E(u_{t-2}^2)} = \sigma^2.$

homoscedasticity

For the second autocovariance coefficient,
$\gamma_2 = cov(y_t, y_{t-2}) = E[y_t - E(y_t)][y_{t-2} - E(y_{t-2})]$

Using the same rules as applied above for the lag 1 covariance

$$
\begin{aligned}
&\gamma_2 = E[(y_t)(y_{t-2})] \\
&= E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + ...)(u_{t-2} + \phi_1 u_{t-3} + \phi_1^2 u_{t-4} + ...)] \\
&= E[(\phi_1 u_{t-2}^2 + \phi_1^4 u_{t-3}^2 + ... + cross - products] \\
&= \phi_1^2 \sigma^2 + \phi_1^4 \sigma^2 + ... \\
&= \phi_1^2 \sigma^2 (1 + \phi_1^2 + \phi_1^4 + ...) \\
&= \frac{\phi_1^2 \sigma^2}{(1 - \phi_1^2})
\end{aligned}
$$

- If these steps were repeated for γ_3, the following expression would be obtained

$$\gamma_3 = \frac{\phi_1^3\sigma^2}{(1-\phi_1^2})$$

and for any lag s, the autocovariance would be given by

$$\gamma_s = \frac{\phi_1^s\sigma^2}{(1-\phi_1^2})$$

The acf can now be obtained by dividing the covariances by the variance:

$$\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{\left(\frac{\phi_1 \sigma^2}{(1-\phi_1^2)}\right)}{\left(\frac{\sigma^2}{(1-\phi_1^2)}\right)} = \phi_1$$

$$\tau_2 = \frac{\gamma_2}{\gamma_0} = \frac{\left(\frac{\phi_1^2 \sigma^2}{(1-\phi_1^2)}\right)}{\left(\frac{\sigma^2}{(1-\phi_1^2)}\right)} = \phi_1^2$$

$$\tau_3 = \phi_1^3$$

...

$\tau_s = \phi_1^s$ which is the autocorrelation at lag s. Thus, $corr(y_t, y_{t-s}) = \phi_1^s$.

THE PARTIAL AUTOCORRELATION FUNCTION (DENOTED τ_{kk})

- Measures the correlation between an observation k periods ago and the current observation, after controlling for observations at intermediate lags (i.e. all lags < k). $Corr.(Y_t, Y_{t-3})$
- So τ_{kk} measures the correlation between y_t and y_{t-k} after removing the effects of $y_{t-k+1}, y_{t-k+2}, \ldots, y_{t-1}$.

  Controlling for Y_{t-2}, Y_{t-1}
- **Example:** the pacf for lag 3 would measure the correlation between y_t and y_{t-3} after controlling for the effects of y_{t-1} and y_{t-2}.

  seasonality, τ_{11}
- At lag 1, the acf = pacf always.

  $Corr(Y_t, Y_{t-2})$
- At lag 2, $\tau_{22} = (\tau_2 - \tau_1^2)/(1 - \tau_1^2)$ where τ_1 and τ_2 are the autocorrelation coefficients at lags 1 and 2 respectively.
- For lags 3+, the formulae are more complex.

  $\tau_{33} = Corr.(Y_t, Y_{t-3})$

The Partial Autocorrelation Function (denoted τ_{kk}) (cont'd)

$\boxed{ARMA(p, q)}$, MA $\rightarrow$ Lags (q) $\rightarrow$ ACF.

- The pacf is useful for telling the difference between an AR process and an ARMA process.

PACF $\rightarrow$ Lags $\rightarrow$ AR(p)

τ_{kk}

- In the case of an AR(p), there are direct connections between y_t and y_{t-s} only for s$\leq$ p.

Corr (Y_t, Y_{t-s})

- So for an AR(p), the theoretical pacf will be zero after lag p.
- **Invertibility Condition:** In case of a MA(q), it can be written as an AR(∞), so there are direct connections between y_t and all its previous values.

MA(1) $\rightarrow$ AR(∞)

MA(q) $\rightarrow$ AR(∞)

- By combining the AR(p) and MA(q) models, we can obtain an ARMA(p,q) model:

$L y_t$ (AR(p)) MA(q)

$y_t =$

$\underbrace{\mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}}_{}+ \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q} + u_t$

with $E(u_t) = 0$; $E(u_t^2) = \sigma^2$; $\boxed{E(u_t u_s) =} 0,\ t \neq s$

or

$L y_t^2$. No Autocorr.

$$\phi(L) y_t = \mu + \theta(L) u_t$$

ACF

PACF.

$|\phi| < 1$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ AR.

and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$ MA.

ARMA(2,2)

(3,2) - - -

Summary of the Behaviour of the acf for AR and MA Processes

An autoregressive process has

- a geometrically decaying acf
- number of spikes of pacf = AR order

AR(p)

p = signif. no of spikes of τ_{kk}.

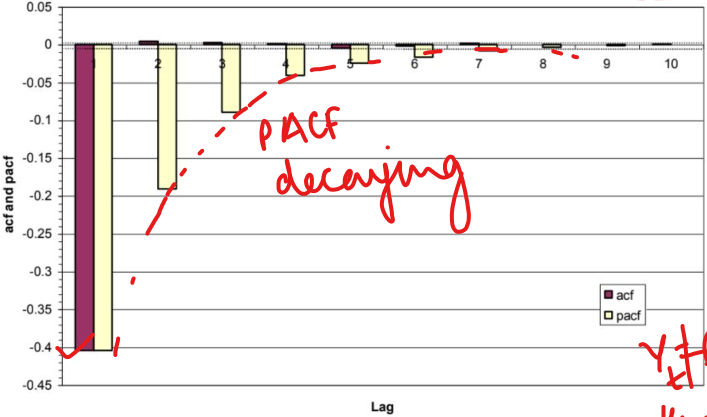
A moving average process has

- a geometrically decaying pacf
- Number of spikes of acf = MA order

MA(q)

The acf and pacf are not produced analytically from the relevant formulae for a model of that type, but rather are estimated using 100,000 simulated observations with disturbances drawn from a normal distribution.

The acf and pacf for a MA(1) Model : $y_t = -0.5u_{t-1} + u_t$



ACF = 1 = PACF

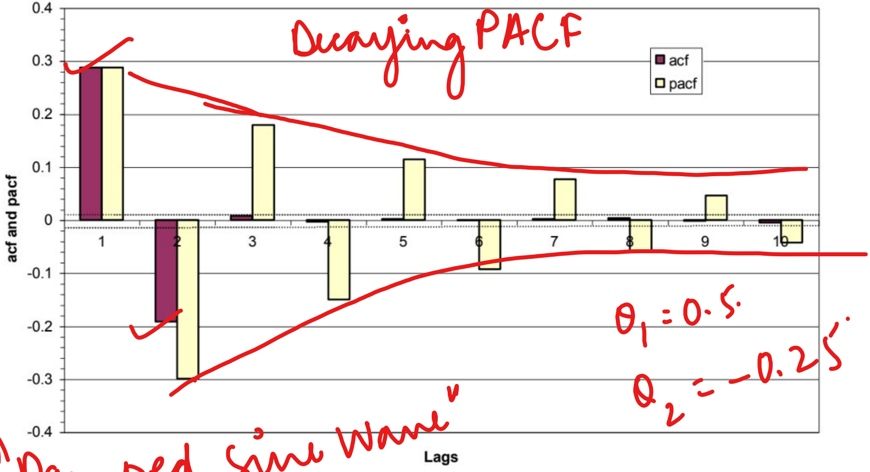
$\tau_k = \tau_{kk}$ at Lag 1

$\tau_1 = \frac{\gamma_1}{\gamma_0} = \tau_{11}$

$y_t \perp\!\!\!\perp (u_{t-2}, u_{t-3}, \ldots)$

ACF and PACF for an MA(2) Model:

$y_t = 0.5 u_{t-1} - 0.25 u_{t-2} + u_t$ MA(2)



$\theta_1 = 0.5$

$\theta_2 = -0.25$

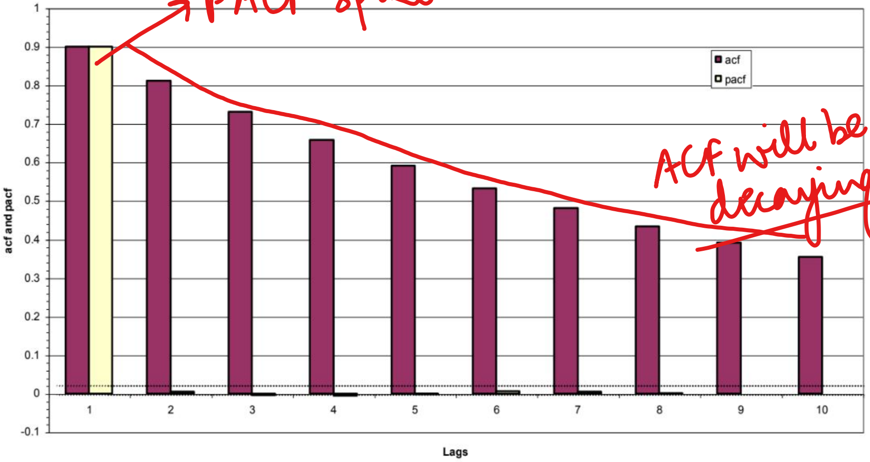
"Damped sine wave"

ACF and PACF for a slowly decaying AR(1) Model:

$y_t = \boxed{0.9} y_{t-1} + u_t$ AR(1)

~1

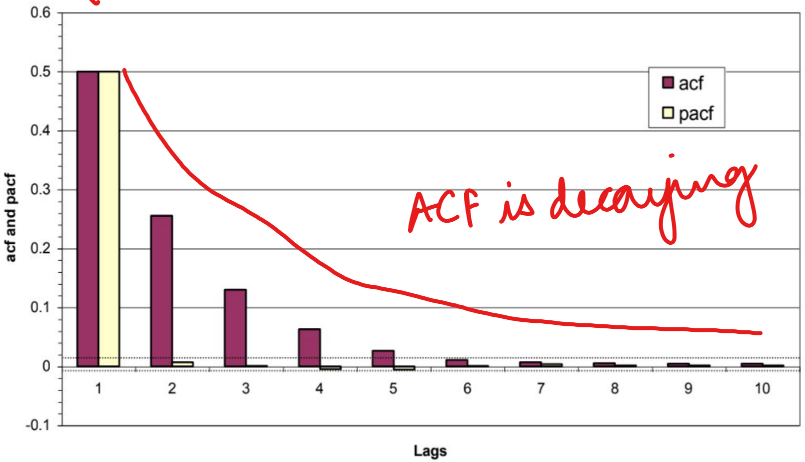
PACF spike



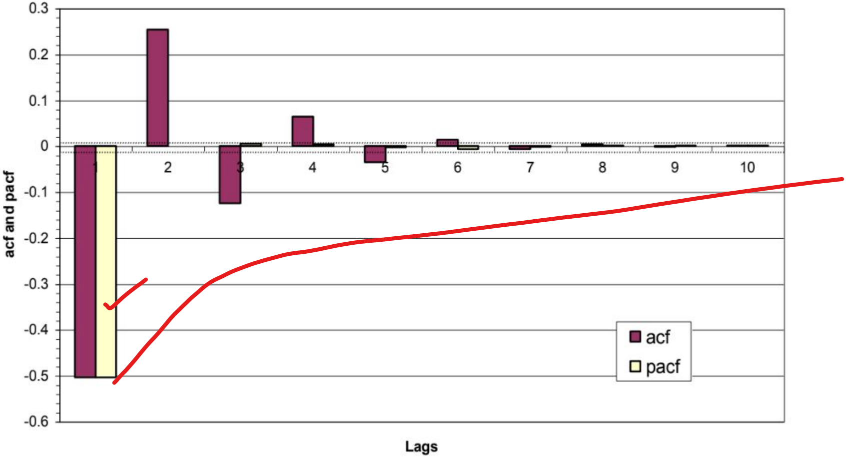
ACF will be decaying

ACF and PACF for a more rapidly decaying AR(1) Model: $y_t = 0.5 y_{t-1} + u_t$

< 1 .

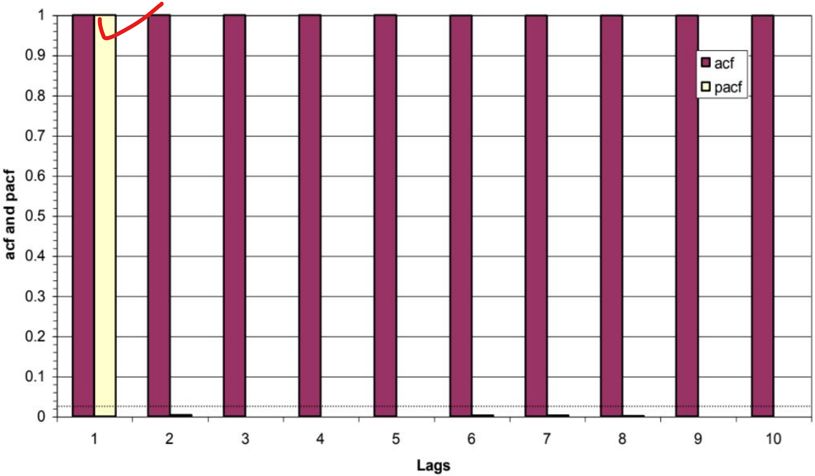


ACF and PACF for a more rapidly decaying AR(1) Model with Negative Coefficient: $y_t = -0.5 y_{t-1} + u_t$



ACF and PACF for a Non-stationary Model (i.e. a unit coefficient): $y_t = y_{t-1} + u_t$

ACF is not Decaying



- Box and Jenkins (1970) were the first to approach the task of estimating an ARMA model in a systematic manner. There are 3 steps to their approach:
  1. Identification
  2. Estimation
  3. Model diagnostic checking

based on ACF & PACF.

Step 1:

- Involves determining the order of the model.
- Use of graphical procedures (plotting data over time, acf and pacf)
- A better procedure is now available

Information criteria AIC, BIC.

ARMA(2,2)
$\hat{\beta}$s.

Step 2:

- Estimation of the parameters of the model specified in step 1.
- Can be done by using least squares or maximum likelihood depending on the model.

Step 3:

- Model checking — Robustness

Box and Jenkins suggest 2 methods :

- deliberate overfitting → Adjusted R^2.
- residual diagnostics

- Identification would typically not be done using acf and pacf plots.
- Main reason is the 'messy' real data as they unfortunately rarely exhibit the above shown simple patterns.
- This gives motivation for using information criteria (IC), which embody 2 factors
  1
  - a term which is a function of the RSS
  2
  - some penalty for the loss of degrees of freedom for adding extra parameters. Adding new variable or an additional lag to a model will have two competing effects on IC - RSS will fall but the value of penalty term will increase.
- The objective is to choose the number of parameters which minimises the information criterion.

Min RSS , RSS↓.

- The information criteria vary according to how stiff the penalty term is.
- The three most popular criteria are Akaike's (1974) information criterion (AIC), Schwarz's (1978) Bayesian information criterion (SBIC), and the Hannan-Quinn criterion (HQIC).

$\hat{\sigma}^2 = \frac{\Sigma(e-\bar{e})^2}{N}$

$$AIC = \ln(\hat{\sigma}^2) + 2k/T$$
$$SBIC = \ln(\hat{\sigma}^2) + \frac{k}{T}\ln(T)$$
$$HQIC = \ln(\hat{\sigma}^2) + \frac{2k}{T}\ln(\ln(T))$$

$T \to \infty$.

$2k/T \to 0$

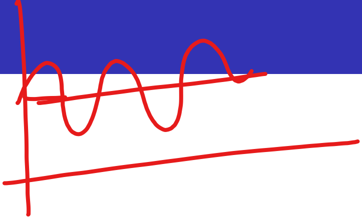
AIC ↓

where $\hat{\sigma}^2$ is the residual variance, $k = p + q + 1$, T = sample size. So we minimized IC s.t $p \leq \overline{p}, q \leq \overline{q}$.
SBIC embodies a stiffer penalty term than AIC.

- Which IC should be preferred if they suggest different model orders?
  - SBIC is strongly consistent but (inefficient). Var. is not min.
  - AIC is not consistent, and will typically pick "bigger" models.

ARIMA Models

stationarity



- As distinct from ARMA models. The I stands for integrated.
- An integrated autoregressive process is one with a characteristic root on the unit circle. non-stationary.
- Typically researchers difference the variable as necessary and then build an ARMA model on those differenced variables.
- An ARMA(p,q) model in the variable differenced d times is equivalent to an ARIMA(p,d,q) model on the original data.

difference only I.

$\approx 1 \rightarrow I(1)$

$y_t \sim I(2)$

$\Delta^2 y_t$.

- **The Box-Jenkins approach step by step**
  - Step 1 Calculate the ACF and PACF of the raw data, and check whether the series is stationary or not. If the series are stationary go to step 3, if not go to step 2.
  - Step 2 Take the logarithm and the first differences of the raw data and calculate the ACF and PACF for the first logarithmic differenced series.
  - Step 3 Examine the graphs of the ACF and PACF and determine which models would be good starting points.
  - Step 4 Estimate those models.

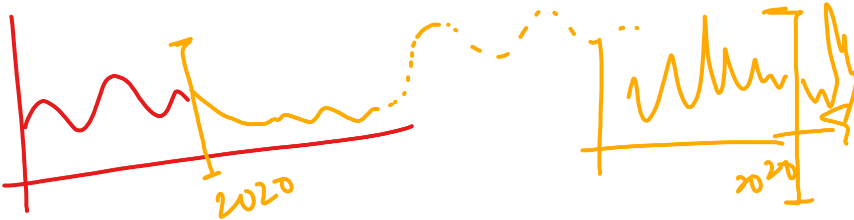
(2,2)
or (3,3)
AIC, BIC.

$Y_{t-1}, Y_{t-2}, Y_{t-3}$

- • **Step 5 For each of these estimated models:**
  - check to see if the parameter of the longest lag is significant. If not, then you probably have too many parameters, and should decrease the order of p and/or q.
  - check the ACF,and PACF of the errors. If the model has at least enough parameters, then all error ACFs and PACFs will be insignificant.
  - check the AIC and SBC together with the adj-Rsquare of the estimated models to detect which model is the parsimonious one (i.e. the one that minimizes AIC and SBC and has the highest adj-Rsquare).
- **Step 6 If changes in the original model are needed, go back to step 4.**

Aim is to Min RSS.

- But why tests for non-stationarity are necessary?
  - Stationarity of a series can strongly influence its behaviour and properties. Like, for a stationary series 'shocks' to the system will gradually die away.
  - But, this can be contrasted with the case of non-stationary data, where the persistence of shocks will always be infinite.



Two types of Non-stationarity?

- Various definitions of non-stationarity exist.
- There are two models which have been frequently used to characterise non-stationarity: the random walk model with drift:

$$y_t = \mu + y_{t-1} + u_t$$

$= 1$.

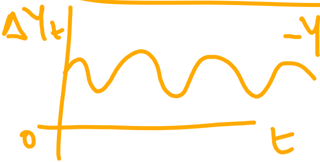
and the deterministic trend process:

$$y_t = \alpha + \beta t + u_t$$

where u_t is iid in both case.

$\Delta Y_t = Y_t - Y_{t-1}$

- Consider again the simplest stochastic trend model:



$-Y_{t-1}$

$$y_t = 1\, y_{t-1} + u_t$$

$Y_t \sim I(2)$

$\Delta^2 Y_t$.

$$\Delta y_t = u_t$$

ΔY_t.

- We can generalise this concept to consider the case where the series contains more than one "unit root". That is, we would need to apply this first difference operator, Δ, more than once to induce stationarity.

Definition

If a non-stationary series, y_t must be differenced d times before it becomes stationary, then it is said to be integrated of order d. We write $y_t \sim I(d)$.

$Y_t \sim I(1)$

So if $y_t \sim I(d)$ then $\Delta^d y_t \sim I(0)$. $\rightarrow \Delta Y_t \sim I(0)$.

An I(0) series is a stationary series An I(1) series contains one unit root, e.g $Y_t = y_{t-1} + u_t$

- An I(2) series contains two unit roots and would require differencing twice to induce stationarity.
- I(1) and I(2) series can wander a long way from their mean value and cross this mean value rarely.
- I(0) series should cross the mean frequently.
- The majority of economic and financial series contain a single unit root, although some are stationarity and consumer prices have been argued to have 2 unit roots.

I(2)

How do we test for a unit root?

- The early and pioneering work on testing for a unit root in time series was done by Dickey and Fuller (Dickey and Fuller 1979, Fuller 1976). The basic objective of the test is to test the null hypothesis that $\phi = 1$ in: $-y_{t-1}$ $-y_{t-1}$

$$y_t = \phi y_{t-1} + u_t$$

against the one-sided alternative $\phi < 1$. So we have

H_0 : series contains a unit root (non-stat.)

vs. H_1 :series is stationary.

$$\Delta y_t = (\phi - 1) y_{t-1} + u_t.$$

- We usually use the regression:

$$\Delta y_t = \psi y_{t-1} + u_t$$

- Thus, a test of $\phi = 1$ is equivalent to a test of $\psi = 0$ (since $\phi - 1 = \psi$).

$\phi = 1.$, $\psi = 0.$

$t = \frac{\hat{\beta} - \beta}{SE\hat{\beta}}$, $H_0: \beta = 0$ (Regression)

$\sim$ t-Distn.

$$Test \quad Stat = \frac{\hat{\phi}}{SE\hat{\phi}}$$

$\not\sim$ t and F

$H_0: \phi \not\sim$

- The test statistics does not follow the usual t-distribution under the null, since the null is one of non-stationarity, but rather follows a non-standard distribution.
- Critical values are derived from Monte Carlo experiments in, for example, Fuller (1976). Relevant examples of the distribution are shown in the table below.

Critical Values for the DF Test

Significance Level	10%	5%	1%
C.V for constant but no trend	-2.57	-2.86	-3.43
C.V for constant and trend	-3.12	-3.41	-3.96

- The null hypothesis of a unit root is rejected in favour of the stationary alternative in each case if the test statistics is more negative than the critical value.

Thank you!!